REPORT DOCUMENTATION PAGE

AFRL-SR-AR-TR-05-

0359

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
	Final Report	01 Jan 2002 - 31 Dec 2004

4. TITLE AND SUBTITLE

Data Refinement for Confidence Management in Model-Based Predictions

5a. CONTRACT NUMBER

5b. GRANT NUMBER

F49620-02-1-0068

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Professor Roger Ghanem

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Department of Civil Engineering
The Johns Hopkins University
Baltimore MD 21218

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

USAF/AFRL
AFOSR
875 North Randolph Street
Arlington VA 22203

NA

10. SPONSOR/MONITOR'S ACRONYM(S)

AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution Statement A. Approved for public release; distribution is unlimted.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The activities under this grant were in the general area of Verification and Validation using the Polynomial Chaos formalism developed by the PI over the course of the past 15 years. In particular, three significant questions were formulated and addressed in the course of this research:

1. How to develop polynomial chaos representations of physical parameters from experimental measurementss of these parameters?
2. How to propagate the uncertainty in these parameters (as reflected in their Polynomial Chaos representations) into the dynamical behavior of the physical system.
3. How important is data refinement: what is the significance, on the predictive value of a computational model, of collecting additional experimental measurements as opposed to performing further numerical refinement (using, for example, mesh refinement). The remainder of this report will review the highlights of each of the above three topics

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	10	

Final Report: Data Refinement for Confidence Management in Model-Based Predictions

Roger Ghanem
University of Southern California, CA, USA

AFOSR Grant F49620-02-1-0068

1 Summary of Developments

This report documents activities completed under AFOSR Grant F49620-02-1-0068 from January 2002-June 2005. Although activities under this grant were completed in December 2004, work towards its original objectives was only recently completed on June 30th 2005. In particular, some of the research directions that were initially pursued could only be completed on that date. Three graduate students were each partially supported on this grant, one of whom (Dr. Debraj Ghosh) successfully defended his doctoral thesis, at the Johns Hopkins University, on June 30 2005, while the other two students (Mr. Sonjoy Das and Mr. Alireza Doostan) have already completed significant work towards their dissertations. Their work on this AFOSR portion of their work will culminate in a MS thesis for each of them.

The activities under this grant were in the general area of Verification and Validation using the Polynomial Chaos formalism developed by the PI over the course of the past 15 years. In particular, three significant questions were formulated and addressed in the course of this research:

1. How to develop polynomial chaos representations of physical parameters from experimental measurements of these parameters ? This provides the link between experimental evidence and the stochastic predictive process. It has enabled us to eliminate severe assumptions associated with probabilistic approaches in general, namely having to invoke various assumptions as to the probabilistic measure of the data (gaussian, lognormal, various Askey scheme measures, ...)

2. How to propagate the uncertainty in these parameters (as reflected in their Polynomial Chaos representations) into the dynamical behavior of the physical system. Specifically, a novel approach is developed to the random eigenvalue problem using Polynomial Chaos representations coupled with Stochastic

Galerkin Projections. The new approach provides both efficient algorithms for characterizing the solution to the random eigenvalue problem as well as new insight as to the significance and usefulness of this solution.

3. How important is data refinement: what is the significance, on the predictive value of a computational model, of collecting additional experimental measurements as opposed to performing further numerical refinement (using, for example, mesh refinement). This "data refinement" issue is significant if predictive tools are to become surrogates to physical experiments.

The remainder of this report will review the highlights of each of the above three topics. Details of the analysis are contained in seven papers which are various stages of the publication process: one is in press for the AIAA Journal [24], one has been accepted for publication in the International Journal of Numerical Methods in Engineering [7], one is under review for the IJNME [18, 23], one is under review for the Journal of Computational Physics [17], and two are still in preparation [8, 15].

2 Characterization of Chaos Representations from Measurements

In a series of previous publications, the PI presented a mathematical framework for the characterization and propagation of uncertainty in physical systems [31, 16, 13, 12, 14, 27, 19, 22, 21, 20, 29, 28, 6, 5, 24]. That work is based on the adaptation of multiple Wiener integral representations [33, 4] to finite-dimensional spaces and their implementation into a weighted residual scheme for the stochastic characterization of the solution of stochastic partial differential equations. The restriction of the representations to finite-dimensional uncertainty (i.e. stochastic processes characterized by a finite-dimensional random vector) permits the generalization of the Wiener constructions which had used polynomials orthogonal with respect to Gaussian measure. Extensions of that work to the Askey scheme were recently developed [35, 34] together with extensions using non-orthogonal representations [1, 2] and representations in terms of wavelets [25, 26]. While the above extensions are limited to representations in terms of independent random variables, extensions using finite-dimensional dependent random vectors (such as appearing, for instance, in the finite-dimensional Karhunen-Loeve representation of an arbitrary stochastic process) have also been completed [30]. These so-called polynomial chaos expansions, coupled with stochastic projection mechanisms, provide a general method for characterizing the solution of problems of mathematical physics whose parameters have been described as stochastic process. An analysis of the error associated with this method has already been developed [3, 2]. The mathematical foundation of these developments, which permits their rigorous mathematical analysis, lies in functional analysis and in the observation that second-order random vectors (i.e. loosely speaking, those with finite second order statistics) form a Hilbert space.

Implicit in all these developments is the assumption that the parameters in the governing equations have been accurately characterized as stochastic processes. In this case, the formalism described above, based on the Chaos expansions, can be viewed as an efficient procedure to propagate the probabilistic measures from the system parameters to the solution, and as such appears to be an alternative to approaches based on Monte Carlo sampling or perturbation expansions [36]. As noted above, it is often the case that system parameters are not known with enough resolution to permit their accurate characterization as stochastic processes. In such cases, the approach based on Chaos developments provides a unique perspective on the problem as the sensitivity of the system parameters to additional information can be cast as perturbing their coordinates with respect to the polynomial chaos, which in turn is readily described as perturbations in Hilbert space. The impact of these perturbations on the chaos coordinates of the solution to the governing equations can then be viewed as quantifying the impact of refining the probabilistic measure of the data on the predictive capability of the mathematical model. Ours efforts in this direction have relied on the maximum-likelihood arguments to compute estimates of the Chaos coordinates of parameters from associated statistical samples [7]. The same problem has also been addressed using the Bayesian framework for parameter estimation. The benefit of this approach lies in its ability to characterize the statistics of the estimates, thus enabling the determination of their accuracy and their sensitivity to further data. In addition, a propagation of the error associated with these estimates to the predictions of the stochastic governing equations has been developed. This clearly provides an essential ingredient for any model validation process.

The procedure described in this section is described in various recent publications by PI and co-workers [9, 17, 7]

3 Data Refinement

By representing the coefficients in the Chaos representations as random variables themselves, it becomes possible to determine the influence of the uncertainty in their value on the predicted solution.

Specifically, as experimental data is obtained, estimates are updated for any statistics of the stochastic process being represented. In particular, the chaos coefficients are a particular kind of statistics. When a stochastic representation is available for these coefficients, reflecting the uncertainty in their value, the impact of this uncertainty can be propagated to the solution, and the worth of additional information can be ascertained. It is assumed that the uncertainty in the chaos coefficients is independent of the uncertainty in the mechanistic parameters themselves. This results in an increase in the dimension of the chaos in which the solution is described.

The problem then becomes a standard problem in stochastic computational mechanics, and can be very efficiently solved using standard procedures established by the PI for Stochastic Galerkin Projections. It should be noted here that the so-called

"generalized chaos" expansions in terms of non-gaussian measures cannot be applied to this task without significant assumptions, specifically in terms of completeness in infinite dimensional spaces (in this case, the tensorized construction of bases using one-dimensional spaces as building blocks only works for the Gaussian and Poisson measures [32, 10, 11]).

The procedure described in this section is described in an invited paper to the Journal of Computational Physics, and has been presented in conference proceedings [9, 8, 17]

4 Random Eigenvalue Problem: Polynomial Chaos Representation

The polynomial chaos decomposition involves representing a stochastic process with respect to a Hilbertian basis consisting of the multidimensional Hermite polynomials of orthogonal gaussian variables. For the random eigenvalue problems where the stiffness K is a function of a set of random variables $\boldsymbol{\xi}$, the polynomial chaos representation of the eigenvectors and eigenvalue has also been used. Thus starting with

$$K(\boldsymbol{\xi})\phi = \lambda M \phi \tag{1}$$

The eigenvalues and the eigenvectors are then expressed as

$$\phi = \sum_{i=0}^{P} \psi_i \phi_i, \qquad \lambda = \sum_{i=0}^{P} \psi_i \lambda_i$$

where ψ_i are zero-mean, orthogonal polynomials in $\boldsymbol{\xi}$ such that,

$$\psi_0 \equiv 1; \qquad \langle \psi_i \rangle = 0 \quad i > 0; \qquad \langle \psi_i \psi_j \rangle = \delta_{ij} \langle \psi_i^2 \rangle \qquad i, j > 0.$$

ϕ_i and λ_i, the coefficients in the expansion, are calculated using their generalized Fourier coefficient expressions,

$$\phi_i = \frac{\langle \phi \psi_i \rangle}{\langle \psi_i^2 \rangle}, \qquad \lambda_i = \frac{\langle \lambda \psi_i \rangle}{\langle \psi_i^2 \rangle} \; .$$

The denominator in the above expressions can be exactly evaluated, whereas the numerator must be estimated, for example, by Monte Carlo sampling. In this case, each realization of the set of random variables $\boldsymbol{\xi}$ is associated with a specific realization of the stochastic system and the corresponding eigenproblem is solved for λ and ϕ. The realization of ψ_i is simultaneously computed as a polynomial form in the set $\{\xi_i\}$. The coefficients ϕ_i and λ_i are then estimated by interpreting the inner product as a statistical average estimated using the realizations. Once the chaos coefficients have been computed, realizations of the eigenvalues or eigenvectors can

be readily synthesized and used to estimate higher order statistics of the probability density functions. In particular, the first and second order statistical moments of the eigenvalues are obtained as,

$$\bar{\lambda} \equiv \langle \lambda \rangle = \lambda_0, \qquad Var(\lambda) = \sum_{i=1}^{P} \langle \psi_i^2 \rangle \lambda_i^2$$

Similarly, for the eigenvectors,

$$\bar{\phi} \equiv \langle \phi \rangle = \phi_0 \ .$$

It can be noted that the polynomial chaos approach does not require explicit knowledge of the derivatives of the stiffness and mass matrices with respect to the random parameters.

The procedure described in this section has been presented in journal publications [23, 18] and various conference proceedings.

4.1 Intrusive, Galerkin-Based Approach

An alternative way to compute the coefficients in the chaos representation of the eigensolution, which does not require Monte Carlo sampling, is to pose the problem as an optimization problem. In particular, the chaos coefficients of eigenvalues and eigenvectors are sought such that the error between the left and right sides of the eigen-equation is minimized. This is described next.

Consider the random eigenvalue problem given by the equation

$$K(\boldsymbol{\xi})\phi = \lambda\phi \ . \tag{2}$$

Let also expand K in polynomial chaos expansion,

$$K = \sum_{i=0}^{L-1} \psi_i K_i \ . \tag{3}$$

Note that an L-term expansion is used for the matrix K representing the data, and a P-term expansion is used for the unknown eigenvectors and the eigenvalues. Now onward, the subscript l is removed from λ snd ϕ for convenience, implicitly assuming that the rest of the formulation is done for a single mode. Let us define Substituting Chaos expansions into the eigen-equation results in

$$\sum_{i=0}^{L-1}\sum_{j=0}^{P-1} \psi_i\psi_j K_i\phi^{(j)} = \sum_{i=0}^{P-1}\sum_{j=0}^{P-1} \psi_i\psi_j \lambda^{(i)}\phi^{(j)} \ . \tag{4}$$

The above equality is next interpreted to hold in the weak sense. Thus multiplying the equation by ψ_k and taking the inner product in the Hilbert space of second order random variables results in,

$$\sum_{i=0}^{L-1}\sum_{j=0}^{P-1} <\psi_i\psi_j\psi_k> K_i\phi^{(j)} = \sum_{i=0}^{P-1}\sum_{j=0}^{P-1} <\psi_i\psi_j\psi_k> \lambda^{(i)}\phi^{(j)} \ , \quad k = 0, \ldots, P-1 \ , \tag{5}$$

that can be rewritten as,

$$\mathbf{K}\Phi = \mathbf{\Lambda}\Phi \ . \tag{6}$$

Here Φ is nP dimensional vector, constructed by arranging P number of n dimensional vectors $\phi^{(j)}$ in a single column. $\mathbf{K}$ and $\mathbf{\Lambda}$ are the matrices of dimension $(nP \times nP)$, described below. set of vectors and scalars describe the behavior of a single physical mode only.

A Newton method is used to solve the associted optimization problem.

It should be noted that in the Monte Carlo simulation, and in the existing simulation based chaos expansion method, the eigenvalues are ordered in ascending or descending order according to their numerical values in each realization. For repeated eigenvalue cases and some closely spaced eigenvalue cases this ordering strategy results in modal switching whereby the relative ordering changes for two modes associated with two distinct behavior. The closeness of the modes are measured with respect to the strength of perturbation. This Galerkin-based technique deals with statistical description of both the eigenvalues and eigenvectors simultaneously, resulting in eigenpair ordering rather than the eigenvalue ordering. These two ordering techniques differ in the above mentioned closely spaced or repeated eigenvalues cases. In such cases, studying the behavior of the modal subspace could be more useful rather than considering the individual modes. However, this paper considers only the widely spaced modes where such situations does not arise.

The non-intrusive approach is described in detail in [24].

4.2 Application to large scale system

Implementing the present concepts in the context of large-scale systems (millions of degrees of freedom) presents a number of serious challenges. A number of issues have been encountered and resolved.

1. The size of the jacobian arising in the Newton's method is very large, thus using the direct linear solvers become prohibitive, an iterative solver is needed. For a small system, it is observed that the MINRES algorithm is a very efficient choice.

2. Mat-vec multiplication: Due to the large size of the jacobian, it can not be stored, thus all the rows are recomputed whenever a matrix-vector product is sought. This results in heavy computational burden. More efficient methods for the mat-vec product are needed, this is a crucial hinge.

4.3 Repeated and Closely Spaced Modes

If the eigenvalues are clustered, then the corresponding eigenvalues and the eigenvectors are highly sensitive to the system parameters. In the chaos expansion of such eigenvalues and the eigenvectors, two problems are observed. First, neither the non-intrusive nor the intrusive methods can adequately capture the variability of the eigensolution. Second, the results obtained using these two methods are quite different. But both of them can capture the stable invariant subspace formed by the set of the eigenvectors corresponding to the cluster of the eigenvalues. An analysis of the problem can be done using a simple mass-spring system.

Recently an enriched version of the polynomial chaos expansion has been proposed. The idea is to simply augment the set of basis functions with some special functions that anticipate particular behaviors in the response of the system. These functions will typically consist of a step function, an absolute-value function, or a rational polynomial. For a system with one random parameter, it is observed that for the expansion of the eigenvalue, adding an absolute-value function and for the expansion of the eigenvector, adding a step function improves the result significantly.

Once the Chaos representation has been computed, a corresponding large statistical sample can be easily generated, from which estimates can be developed for the probability density function of the response quantity of interest.

Results from this part of the work have been presented at various conferences and submitted for publication in a refereed journal [23].

5 On-Going Work

Dr. Ghosh is spending the next two months completing the implementation of the stochastic eigen-analysis to large scale and arbitrary systems. He is supported on this work by a discretionary account available to the PI at USC.

Mr. Doostan will be finishing his Doctoral thesis within the next 15 months. He is working on integrating the errors from data limitations, numerical truncation and statistical sample limitations in order to formulate a comprehensive error budget that can be used in defining a rigorous, quantifiable and constructive validation process. Mr. Doostan's research is supported by a recent NSF grant.

Mr. Das will be also be finishing his doctoral thesis within the next 15 months. He is working on alternative procedures for polynomial chaos characterization.

References

[1] I. Babuška and P. Chatzipantelidis. On solving elliptic stochastic partial differential equations. *Computer Methods in Applied Mechanics and Engineering*, 191(37-38):4093–4122, 2002.

[2] I. Babuška, R. Tempone, and G. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM Journal on Numerical Analysis*, 42(2):800–825, 2004.

[3] F. Benth and J. Gjerde. Convergence rates for finite element approximations of stochastic partial differential equations. *Stochastics and Stochastic Reports*, 63:313–326, 1998.

[4] R.H. Cameron and W.T. Martin. The orthogonal development of nonlinear functionals in series of fourier-hermite functionals. *Ann. Math.*, 48:385–392, 1947.

[5] B. Debusschere, H. Najm, A. Matta, O. Knio, R. Ghanem, and O. Le Maitre. Protein labeling reactions in electrochemical microchannel flow: Numerical simulation and uncertainty propagation. *Physics of Fluids*, 15(8):2238–2250, 2003.

[6] B. Debusschere, H. Najm, P. Pebay, O. Knio, R. Ghanem, and O. Le Maitre. Numerical challenges in the use of polynomial chaos representations for stochastic processes. *SIAM Journal of Scientific Computing*, 26(2):698–719, 2004.

[7] C. Descelliers, R. Ghanem, and C. Soize. Maximum likelihood estimation of stochastic chaos representation from experimental data. *International Journal of Numerical Methods in Engineering*, to appear, 2005.

[8] A. Doostan and R. Ghanem. An stochastic error analysis of the polynomial chaos expansions. *Computer Methods in Applied Mechanics and Engineering*, In preparation: Invited paper.

[9] A. Doostan and R. Ghanem. Characterization of stochastic system parameters from experimental data: A bayesian inference approach. In *Proceedings of ICOSSAR-05*, June 2005.

[10] N. Dunford and J. Schwartz. *Linear Operators: Part I, General Theory, page 402.* Wiley in 1988, 1957.

[11] K.O. Friedrichs and H.N. Shapiro. Integration on functionals. Technical report, Courant Institute of Mathematical Sciences, 1957.

[12] R. Ghanem. Probabilistic characterization of transport in heterogeneous porous media. *Computer Methods in Applied Mechanics and Engineering*, 158(3-4), 1998.

[13] R. Ghanem. Scales of fluctuation and the propagation of uncertainty in random porous media. *Water Resources Research*, 34(9):2123–2136, 1998.

[14] R. Ghanem. Ingredients for a general purpose stochastic finite elements formulation. *Computer Methods in Applied Mechanics and Engineering*, 168(1-4):19–34, 1999.

[15] R. Ghanem and S. Das. Maximum entropy and fisher information for the estimation of polynomial chaos representations from experimental data. *Computer Methods in Applied Mechanics and Engineering*, In preparation.

[16] R. Ghanem and S. Dham. Stochastic finite element analysis for multiphase flow in heterogeneous porous media. *Transport in Porous Media*, 32:239–262, 1998.

[17] R. Ghanem and A. Doostan. Characterization of stochastic system parameters from experimental data: A bayesian inference approach. *Journal of Computational Physics*, submitted, 2005.

[18] R. Ghanem and D. Ghosh. A polynomial chaos characterization of the random eigenproblem. *International Journal for Numerical Methods in Engineering*, submitted, 2005.

[19] R. Ghanem and A. Sarkar. Mid-frequency structural dynamics with parameter uncertainty. *Computer Methods in Applied Mechanics and Engineering*, 191:5499–5513, 2002.

[20] R. Ghanem and A. Sarkar. Reduced models for the medium-frequency dynamics of stochastic systems. *JASA*, 113(2):834–846, 2003.

[21] R. Ghanem and P. Spanos. *Stochastic Finite Elements: A Spectral Approach.* Dover, 2002.

[22] D. Ghiocel and R. Ghanem. Stochastic finite element analysis of seismic soil-structure interaction. *ASCE, Journal of Engineering Mechanics*, 128(1):66–77, 2002.

[23] D. Ghosh and R. Ghanem. Basis enrichment in polynomial chaos representations. *International Journal for Numerical Methods in Engineering*, submitted, 2005.

[24] D. Ghosh, R. Ghanem, and J. Red-Horse. Analysis of eigenvalues and modal interaction of stochastic systems. *AIAA Journal*, page to appear, 2005.

[25] O.P. Le Maitre, H. Najm, R. Ghanem, and O. Knio. Multi-resolution analysis of Wiener-type uncertainty propagation schemes. *J. Comp. Phys.*, 197(2):502–531, 2004.

[26] O.P. Le Maitre, Knio. O., H. Najm, and R. Ghanem. Uncertainty propagation using Wiener-Haar expansions. *J. Comp. Phys.*, 197:28–57, 2004.

[27] O. Le Maitre, Knio, M. Reagan, H. Najm, and R. Ghanem. A stochastic projection method for fluid flow. i: Basic formulation. *Journal of Computational Physics*, 173:481–511, 2001.

[28] M. Reagan, H. Najm, B. Debusschere, O.P. Le Maitre, O. Knio, and R. Ghanem. Spectral stochastic uncertainty quantification in chemical systems. *Combustion Theory and Modelling*, 8:607–632, 2004.

[29] S. Sakamoto and R. Ghanem. Polynomial chaos decomposition for the simulation of non-gaussian non-stationary stochastic processes. *ASCE Journal of Engineering Mechanics*, 128(2):190–201, 2002.

[30] C. Soize and R. Ghanem. Physical systems with random uncertainties: Chaos representations with arbitrary probability measure. *SIAM Journal of Scientific Computing*, 26(2):395–410, 2005.

[31] P. Spanos and R. Ghanem. Stochastic finite element expansion for random media. *ASCE Journal of Engineering Mechanics*, 115(5):1035–1053, May 1989.

[32] N. Wiener. Differential space. *J. Math. & Phys.*, 58:131–174, 1923.

[33] N. Wiener. The homogeneous chaos. *Amer. J. Math*, 60:897–936, 1938.

[34] D. Xiu and G. Karniadakis. Modeling uncertainty in flow simulations via generalized polynomial chaos. *J. Comp. Phys.*, 187(1):137–167, 2003.

[35] D. Xiu and G.M. Karniadakis. The wiener-askey polynomial chaos for stochastic differential equations. *SIAM Joural on Scientific Computin*, 24(2):619–644, 2002.

[36] D. Zhang and Z. Lu. An efficient, high-order perturbation approach for flow in random porous media via karhunen-loeve and polynomial expansions. *J. Comp. Phys.*, 194(2):773–794, 2004.